

Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study

Francisco Peñagaricano^{1*}, Kent A. Weigel², Guilherme J. M. Rosa^{1,3*} and Hasan Khatib^{1*}

¹ Department of Animal Sciences, University of Wisconsin, Madison, WI, USA

² Department of Dairy Science, University of Wisconsin, Madison, WI, USA

³ Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA

Edited by:

Peter Dovc, University of Ljubljana, Slovenia

Reviewed by:

Ino Curik, University of Zagreb, Croatia

Anna Tomas Sangenis, Fundacio Investigacio Sanitaria Illes Balears, Spain

*Correspondence:

Francisco Peñagaricano,
Guilherme J. M. Rosa and Hasan
Khatib, Department of Animal
Sciences, University of
Wisconsin-Madison, 1675
Observatory Drive, Madison, WI
53706, USA.
e-mail: fpenagarican@wisc.edu;
grosa@wisc.edu; hkhatib@wisc.edu

Whole-genome association studies typically focus on genetic markers with the strongest evidence of association. However, single markers often explain only a small component of the genetic variance and hence offer a limited understanding of the trait under study. As such, the objective of this study was to perform a pathway-based association analysis in Holstein dairy cattle in order to identify relevant pathways involved in bull fertility. The results of a single-marker association analysis, using 1,755 bulls with sire conception rate data and genotypes for 38,650 single nucleotide polymorphisms (SNPs), were used in this study. A total of 16,819 annotated genes, including 2,767 significantly associated with bull fertility, were used to interrogate a total of 662 Gene Ontology (GO) terms and 248 InterPro (IP) entries using a test of proportions based on the cumulative hypergeometric distribution. After multiple-testing correction, 20 GO categories and one IP entry showed significant overrepresentation of genes statistically associated with bull fertility. Several of these functional categories such as small GTPases mediated signal transduction, neurogenesis, calcium ion binding, and cytoskeleton are known to be involved in biological processes closely related to male fertility. These results could provide insight into the genetic architecture of this complex trait in dairy cattle. In addition, this study shows that quantitative trait pathways inferred from single-marker analyses could enhance our interpretations of the results of genome-wide association studies.

Keywords: gene ontology, gene-set analysis, genome-wide study, male fertility, quantitative trait pathway

INTRODUCTION

The decline in reproductive performance in dairy cattle is a serious concern of farmers and the dairy industry worldwide (Lucy, 2001; Pryce et al., 2004). Most reproductive studies in cattle have focused on fertility of the cow, while male fertility has received much less consideration. Nonetheless, some studies have reported that a significant percentage of reproductive failure in dairy cattle is attributable to male subfertility (DeJarnette et al., 2004). Consequently, bull fertility should not be ignored in breeding schemes aimed at improving the reproductive performance of dairy cattle (Braundmeier and Miller, 2001).

A growing body of evidence suggests that male fertility is influenced by genetic factors. For instance, Khatib et al. (2010) reported that polymorphisms in the *FGF2* and *STAT5A* genes are associated with estimated relative conception rate in Holstein bulls. In addition, Li et al. (2012) showed that *MAP1B* and *PPP1R11*, highly conserved spermatogenesis genes, are significantly associated with sire conception rate (SCR) in the Holstein breed. Recently, some studies have addressed bull fertility using a genome-wide association approach. In this sense, Blaschek et al. (2011) identified several genomic regions associated with non-compensatory fertility in Holstein bulls. Fortes et al. (2012a) reported that genomic regions in BTA2, BTA14, and chromosome X are associated with testicular development, sperm quality, and different hormone levels measured throughout puberty in Brahman bulls.

Furthermore, Peñagaricano et al. (2012) have identified a panel of single nucleotide polymorphisms (SNPs) associated with SCR in a population of 1,755 Holstein sires. Some of the significant SNPs were located within or near genes affecting male fertility, such as the sperm acrosome reaction, chromatin remodeling during spermatogenesis, or the meiotic process during male germ cell maturation (Peñagaricano et al., 2012). Altogether, these studies have identified genomic regions and individual genes associated with bull fertility.

However, although whole-genome association studies have been successful in detecting genetic variants associated with complex traits such as bull fertility, they identify only the most significant genetic markers because of the stringent statistical criteria necessary for minimizing false positive results. Therefore, the identified SNPs represent only a small fraction of the variants contributing to the complex trait and hence the vast majority of genetic variation remains hidden. Thus, testing only for associations with single SNPs is insufficient to dissect the complex genetic structure of a quantitative trait (Peng et al., 2010).

It is increasingly accepted that pathway-based analysis, an alternative strategy based on testing the association of modules of functionally related genes, can greatly complement single-marker association analyses in understanding genetic determinants and biological pathways affecting complex traits (Wang et al., 2007; Chasman, 2008). In a pathway-based analysis, multiple

contributing factors are considered simultaneously, rather than focusing on single markers with the strongest evidence of association with a trait. Therefore, this approach provides a unique opportunity to detect causal pathways and genetic mechanisms underlying complex phenotypes (Weng et al., 2011).

The objective of the present study was to identify biologically relevant pathways affecting SCR as a measure of bull fertility using results of a single-marker association analysis. The identification of fertility pathways would provide better understanding of the genetic architecture of this complex trait in cattle.

MATERIALS AND METHODS

In a previous study, we carried out a genome-wide association analysis for SCR in Holstein dairy cattle (Peñagaricano et al., 2012). Here, we used the results of that single-marker association analysis to perform a pathway-based association analysis. The phenotypic and genotypic data and the statistical models used for the single-marker association analysis were described in detail by Peñagaricano et al. (2012) and are reviewed briefly here.

PHENOTYPIC AND GENOTYPIC DATA

A total of 1,755 Holstein bulls with SCR data, a phenotypic evaluation of bull fertility provided by the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD, USA), were used in the study. In brief, SCR is the expected difference in conception rate of a sire compared with the mean of all other evaluated sires. SCR values from the 1,755 bulls ranged from -10.66 to $+6.80\%$, and the number of breedings per bull ranged from 303 to 111,402. SCR data were obtained from seven consecutive evaluations between August 2008 and December 2010. For bulls with multiple evaluations, the most recent SCR evaluation was used in the analysis. Genotypes of the 1,755 bulls for the Illumina BovineSNP50 Bead Chip were also provided by USDA-ARS Animal Improvement Programs Laboratory. SNPs with minor allele frequencies below 5% were removed. After data editing, 38,650 markers were available.

SINGLE-MARKER ASSOCIATION ANALYSIS

Associations between SNP genotype and SCR were analyzed by using the SNP allele count as a linear covariate, as well as by using the SNP genotype as a categorical variable. The following mixed model was used to analyze SNP genotype as a linear covariate:

$$SCR_{ijkl} = \mu + EVAL_j + \beta_k SNP_{ik} + sire_l + e_{ijkl}$$

where μ is the general mean, $EVAL_j$ is the fixed effect of the j^{th} AIPL-USDA SCR evaluation ($j = 1, 2, \dots, 7$), SNP_{ik} is the number of copies of one allele of the SNP (corresponding to 0, 1, or 2 copies) carried by the i^{th} bull ($i = 1, 2, \dots, 1755$), β_k is the regression coefficient for the SNP considered (also known as the allele substitution effect), $sire_l$ represents the random additive genetic effect of the l^{th} sire ($l = 1, 2, \dots, 222$) of the i^{th} bull, and e_{ijkl} represents the random residual for each observation. For the analysis using SNP genotype as a categorical variable, the same statistical model was used replacing $\beta_k SNP_{ik}$ with SNP_{ik} where SNP_{ik} is the fixed effect of the k^{th} genotypic class ($k = 1, 2, 3$) for the SNP considered carried by the i^{th} bull.

The distributions of $sire_l$ and e_{ijkl} were assumed as $N(0, A\sigma_s^2)$ and $N(0, W^{-1}\sigma_e^2)$, respectively, where A represents the matrix of additive relationships ($1,510 \times 1,510$) between sires in the pedigree and W is a diagonal matrix of order 1,755 with its elements representing reliabilities of the SCR values. Associations between SNPs and SCR were tested using a likelihood ratio test, by comparing the aforementioned models to a reduced model without the SNP effect. The analyses were performed using the *pedigreeemm* package (Vazquez et al., 2010) of the R language/environment (R Development Core Team, 2009).

ASSIGNMENT OF SNPs TO GENES

The UMD 3.1 bovine genome sequence assembly (Zimin et al., 2009) was used for SNP assignment. SNPs were assigned to genes if they were located within the genomic sequence of an annotated gene (i.e., genomic sequence between the start of the first exon and the end of the last exon) or within 20 kb of the 5' or 3' ends of the first and last exon, respectively. The distance of 20 kb was used in order to capture proximal regulatory and other functional regions that may lie outside, but close to, the gene. Additionally, it has been reported that strong linkage disequilibrium can extend up to 20–30 kb in Holstein cattle (de Roos et al., 2008; Kim and Kirkpatrick, 2009). If a SNP was found to be located within or proximal to more than one gene, all of these genes were included in the subsequent analyses. Finally, a gene was considered as significantly associated with SCR if that gene contained at least one SNP with a nominal P -value ≤ 0.05 .

ASSIGNMENT OF GENES TO PATHWAYS

Gene Ontology (GO) and InterPro (IP) databases were used to define functional sets of genes. The GO database (Ashburner et al., 2000), which is probably the most popular among the initiatives for defining functional classes of genes, assigns biological descriptors (named GO terms) to genes on the basis of the properties of their encoded products. These terms fall into three types: biological process, molecular function, and cellular component. Genes assigned to the same GO term can thus be regarded as members of a category of genes that are more closely related in terms of some aspect of their biology than are random sets of genes. The IP database is an integrated documentation resource of protein families, domains, and functional sites (Mulder et al., 2003). Applications of the IP database span a range of biologically important areas that include automatic annotation of protein sequences and genome analysis.

PATHWAY-BASED ASSOCIATION ANALYSIS

The association of a given pathway with SCR was analyzed using a test of proportions based on the cumulative hypergeometric distribution (Tavazoie et al., 1999). This test, also known as Fisher's exact test and commonly used in the analysis of contingency tables, was performed to search for an overrepresentation of significantly associated genes among all the genes in the pathway. The P -value of observing k significant genes in the pathway was calculated by

$$P\text{-value} = 1 - \sum_{i=0}^{k-1} \frac{\binom{S}{i} \binom{N-S}{m-i}}{\binom{N}{m}}$$

where S is the total number of genes that are significantly associated with SCR, N is the total number of genes that are analyzed in the study, and m is the number of genes in the pathway.

To avoid testing overly narrow or broad functional categories, we tested only GO terms with more than 30 genes that were located between levels 5 and 9 in the GO hierarchy. Similarly, we tested only entries in the IP database with more than 30 genes. Furthermore, to account for multiple hypothesis testing (i.e., one for each functional term) and control the false discovery rate, the procedure proposed by Benjamini and Hochberg (1995) was adopted, and the q -value for each test was obtained. The q -value of a test measures the proportion of false positives incurred when that particular test is deemed significant. In this study, functional categories with a q -value ≤ 0.05 were considered significant. All analyses were performed using the procedure FatiGO (Al-Shahrour et al., 2004), implemented on the platform Babelomics (Medina et al., 2010).

RESULTS AND DISCUSSION

In a previous study, we performed a genome-wide association study for bull fertility using a total of 38,650 SNPs spanning the entire bovine genome (Peñagaricano et al., 2012). Associations between SNPs and SCR were analyzed using a mixed linear model that included a random polygenic effect and each SNP genotype, one at a time, as a linear covariate or a categorical variable. The first analysis had more power to detect significant SNPs with additive effects, whereas the second approach allowed the detection of significant markers that show some degree of dominance. Out of the 38,650 genetic markers, 3,353 and 3,096 showed significant associations with SCR (nominal P -value ≤ 0.05) using SNP genotypes as linear covariates or categorical variables, respectively. In total, 4,450 different SNPs showed significant associations with bull fertility in one or both models used for the analysis.

The first step in this pathway-based analysis of a whole-genome association study was to assign SNPs to genes. A total of 21,325 of the 38,650 examined SNPs were located within annotated genes or within 20 kb upstream or downstream from annotated genes. This set of SNPs defined a total of 16,819 genes annotated in the UMD 3.1 bovine genome sequence assembly, which in turn were evaluated for the pathway analysis. In addition, 2,767 of the 16,819 genes had one or more SNPs with nominal P -value ≤ 0.05 ; hence, these genes were defined as significantly associated with bull fertility.

The next step in the analysis was to assign the 16,819 genes into pathways. We tested GO categories and IP entries with more than 30 genes and, in the case of GO terms, located between levels 5 and 9 in the GO hierarchy. A total of 662 GO terms and 248 IP entries met these requirements and hence were tested by a hypergeometric test for enrichment of significant genes associated with bull fertility.

Twenty GO terms showed significant overrepresentation of genes statistically associated with SCR. **Table 1** shows the 11 GO terms that belong to the biological process domain, and **Table 2** displays the nine GO terms classified into the molecular function (4) and cellular component (5) domains, respectively. These tables show the number of genes in each functional category (out of 16,819 genes), the expected number of significant genes under the null hypothesis (i.e., with no significant overrepresentation), and the actual number of significant genes per category. Only GO

terms with q -value ≤ 0.05 were considered significant. In addition, one entry of the IP database, *Haloacid dehalogenase-like hydrolase* (IPR005834) showed significant enrichment (q -value = 0.027) of genes associated with bull fertility. This entry, with a total of 37 genes, showed 16 significant genes, which is 10 more than expected under the null hypothesis. Genes statistically associated with SCR (nominal P -value ≤ 0.05) within each significant category (i.e., the 20 GO terms and IPR005834) are listed in File S1 in Supplementary Material.

Six of the 11 significant GO terms classified into the biological process domain show a very close relationship in the GO hierarchy (**Table 1**). Indeed, *Rho protein signal transduction* (GO:0007266) is part of *Ras protein signal transduction* (GO:0007265) and *Regulation of Ras protein signal transduction* (GO:0046578) is part of *Regulation of small GTPase mediated signal transduction* (GO:0051056). Furthermore, *Ras protein signal transduction* is part of *Small GTPase mediated signal transduction* (GO:0007264), which in turn is part of *Intracellular signaling cascade* (GO:0007242). Interestingly, GO:0046578 and GO:0051056 negatively regulate GO:0007265 and GO:0007264, respectively. Overall, these results point out that an intracellular signaling pathway involving small GTPases and their negative regulators was significantly associated with bull fertility. Interestingly, a recent microarray study showed that *Regulation of small GTPase mediated signal transduction* term was significantly enriched with genes differentially expressed between human sperm that achieved successful fertilization and sperm that failed the intracytoplasmic sperm injection (García-Herrero et al., 2011).

Small GTPases coordinate diverse cellular functions, including actin reorganization, junction dynamics, cell movement, cell cycle, cell transformation, and gene transcription (Bustelo et al., 2007). Several recent studies have shown that small GTPases are relevant molecules for spermatogenesis. For instance, some studies have reported that the translocation of elongating/elongate spermatids across the seminiferous epithelium is assisted by small GTPases present at the site of the ectoplasmic specialization, a unique anchoring junction type located in the testes (Lui et al., 2003). In addition, other studies have shown that small GTPases are involved in regulation of the mammalian acrosome reaction by controlling the membrane fusion system in the sperm (Iida et al., 1999). The acrosome reaction allows spermatozoa to penetrate the zona pellucida and fuse with the oocyte membrane and, hence, this is a crucial step during the fertilization (Brucker and Lipford, 1995). Overall, these results provide further evidence for the possible role of small GTPases and their regulators in spermatogenesis and, hence, in male fertility.

Three significant GO terms classified into the biological process domain are related to neurogenesis (**Table 1**). In fact, *Neuron development* (GO:0048666) is part of *Neuron differentiation* (GO:0030182), which in turn is part of *Neurogenesis* (GO:0022008). One well-known connection between neurons and male fertility is provided by the hypothalamic–pituitary–gonadal axis. Indeed, the maintenance of spermatogenesis depends upon stimulation of the testes by the gonadotropic hormones, follicle-stimulating hormone, and luteinizing hormone, both produced by the pituitary gland in response to gonadotropin-releasing hormone from the hypothalamus (Zirkin, 1998). It is important to

Table 1 | Biological process terms significantly overrepresented with genes statistically associated with bull fertility.

GO ID	Term (GO hierarchy level)	No. genes in the GO term	Expected no. significant genes	Actual no. significant genes	q-Value
0007242	Intracellular signaling cascade (6)	889	146	187	0.032
0007264	Small GTPase mediated signal transduction (7)	352	58	83	0.035
0007265	Ras protein signal transduction (8)	171	28	48	0.032
0007266	Rho protein signal transduction (9)	91	15	27	0.048
0046578	Regulation of Ras protein signal transduction (9)	145	24	39	0.046
0051056	Regulation of small GTPase mediated signal transduction (8)	182	30	47	0.046
0022008	Neurogenesis (6)	325	53	75	0.048
0030182	Neuron differentiation (8)	269	44	65	0.046
0048666	Neuron development (9)	201	33	53	0.034
0032940	Secretion by cell (6)	171	28	45	0.046
0046903	Secretion (5)	197	32	50	0.046

Table 2 | Molecular function and cellular component terms significantly overrepresented with genes statistically associated with bull fertility.

GO ID	Term (GO hierarchy level)	No. genes in the GO term	Expected no. significant genes	Actual no. significant genes	q-Value
MOLECULAR FUNCTION					
0005509	Calcium ion binding (5)	525	86	118	0.018
0016462	Pyrophosphatase activity (6)	578	95	126	0.020
0016818	Hydrolase activity, acting on acid anhydrides (5)	581	96	127	0.018
0017111	Nucleoside-triphosphatase activity (7)	563	93	122	0.022
CELLULAR COMPONENT					
0005856	Cytoskeleton (8)	599	99	124	0.048
0015629	Actin cytoskeleton (9)	185	30	46	0.045
0044420	Extracellular matrix part (5)	65	11	21	0.040
0044430	Cytoskeletal part (9)	373	61	85	0.039
0044459	Plasma membrane part (6)	678	112	138	0.047

note that recent studies have reported *de novo* neurogenesis in postnatal and adult mammalian hypothalamus (Pierce and Xu, 2010). Another well documented connection between neurogenesis and spermatogenesis is the fact that these two biological processes share several key molecular regulators. Heat shock factors and proteins, bone morphogenetic proteins, and inhibitor of DNA-binding/differentiation (Id) proteins, among others, play relevant roles in both processes (Yokota, 2001; Abane and Mezger, 2010; Björk and Sistonen, 2010; Bragdon et al., 2011). Heat shock factors, a family of transcriptional regulators, trigger the expression of genes encoding proteins (i.e., heat shock proteins) that function as molecular chaperones, contributing to establish a cyto-protective state in response to different stress conditions (Björk and Sistonen, 2010). In addition to promoting cell survival under stressful conditions, heat shock factors and proteins are important for developmental processes such as spermatogenesis and neurogenesis (Abane and Mezger, 2010). Bone morphogenetic proteins, potent growth factors belonging to the transforming growth factor beta superfamily, are responsible for many biological processes including embryogenesis, neurogenesis, and spermatogenesis (Bragdon et al., 2011). Id proteins comprise a family of proteins that are positive and negative regulators of proliferation

and differentiation, respectively (Yokota, 2001). Id proteins are key regulators of several developmental processes such as neurogenesis, lymphoid organogenesis, mammary gland development, and spermatogenesis (Yokota, 2001). Altogether, these results present further evidence of the tight connection between neurogenesis and male fertility.

Four GO terms classified into the molecular function domain showed significant association with bull fertility (Table 2). Three of these GO terms showed a very close relationship in the GO hierarchy: *Nucleoside-triphosphatase activity* (GO:0017111) is part of *Pyrophosphatase activity* (GO:0016462), which in turn is part of *Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides* (GO:0016818). Nucleoside-triphosphatases are enzymes that catalyze the hydrolysis of a nucleoside triphosphate in a nucleoside diphosphate and phosphate. Pyrophosphatases are enzymes that catalyze the hydrolysis of a pyrophosphate bond between two phosphate groups. Both types of enzymes, which belong to the family of hydrolases, participate in purine and thiamine metabolism (Zöllner, 1982). Interestingly, Sertoli cells, which are located in the seminiferous tubules and play an essential role in the maintenance and control of spermatogenesis, have shown remarkable activity of these two enzymes

(Casali et al., 2001). In fact, these enzymes play an important role in the physiological control of extracellular levels of nucleotides and nucleosides inside the seminiferous tubules, which in turn can modulate Sertoli cell responses (Casali et al., 2001). Thus, findings of this study provide further support for the relevant role of these enzymes (i.e., nucleoside-triphosphatases and pyrophosphatases) in spermatogenesis.

The *calcium ion binding* (GO:0005509) term showed a significant overrepresentation of genes associated with bull fertility (Table 2). The relationship between calcium and sperm physiology is well documented. Mammalian spermatozoa possess multiple voltage-gated calcium channels and use calcium ion signals to control several physiological responses (Darszon et al., 1999). Indeed, calcium is considered a regulator of sperm motility, a participant in capacitation, and an essential second messenger for the acrosome reaction (Wennemuth et al., 2003). Therefore, our findings provide more evidence of the strong association between calcium and sperm biology and, hence, male fertility.

Three of the five significant GO terms grouped within the cellular component domain are related to the cytoskeleton biology. Indeed, *Actin cytoskeleton* (GO:0015629) and *Cytoskeletal part* are part of *Cytoskeleton* (GO:0005856). The close relationship between cytoskeleton and spermatogenesis is well documented. Sertoli cells are characterized by their well-developed cytoskeleton, which is responsible for the collective organization of the seminiferous epithelium (Mruk and Cheng, 2004). Morphological studies have shown that the Sertoli cell cytoskeleton, among other functions, positions, anchors, and aids in the movement of developing germ cells and participates in the release of mature spermatids from the seminiferous epithelium at spermiation (Mruk and Cheng, 2004). In fact, the different cellular events that occur during spermatogenesis are associated with extensive changes in cell shape and size and germ cell movement, in which the cytoskeleton has a key role (Lie et al., 2010). Interestingly, García-Herrero et al. (2010) have reported that the *Cytoskeleton* term was significantly enriched with genes differentially expressed between sperm of infertile and fertile human males. Therefore, our study provides further evidence of the close relationship between cytoskeletal dynamics and spermatogenesis.

The terms *Secretion by cell* (GO:0032940), *Secretion* (GO:0046903), (Table 1) and *Extracellular matrix part* (GO:0044420), and *Plasma membrane part* (GO:0044459; Table 2) also showed significant enrichment of genes statistically associated with bull fertility. Despite the wide range of functions of these categories, it is worth noting that spermatogenesis and fertilization involve biological processes related to these terms. For instance, many of the functions of the Sertoli cells are carried out by their protein secretions: these cells synthesize and secrete a number of proteins that have functions such as hormone-like or growth factor-like activity, transport functions, enzymatic activities, as well as proteins that contribute to the basement membrane (Griswold, 1998). In addition, it seems that the plasma membrane plays an important role in Sertoli-Sertoli and Sertoli-germ cell interactions (Mruk and Cheng, 2004). Also, it has been reported that the extracellular matrix regulates Sertoli cell differentiation and germ cell development (Hadley et al., 1985).

InterPro is a database that integrates diverse information about families, domains, and functional sites of proteins. One entry of the IP database, *Haloacid dehalogenase-like hydrolase* (IPR:005834), showed significant overrepresentation of genes associated with SCR. IPR:005834 entry contains a group of hydrolase enzymes including L-2-haloacid dehalogenase, epoxide hydrolases, and phosphatases. These enzymes are involved in diverse cellular functions such as membrane transport, signal transduction, metabolism, and DNA repair (Arand et al., 2003). Interestingly, Iguchi et al. (2006), studying gene expression profile in meiotic male germ cells, reported that one haloacid dehalogenase-like hydrolase is highly expressed in the testes. Thus, our finding provides further evidence of the possible involvement of these enzymes in male fertility.

Previous studies have derived gene networks and pathways related to complex traits in livestock species. For instance, Fortes et al. (2012b) have used genome and trait associations, hypothalamic-transcriptome data and transcription factors in order to find candidate genes and gene-gene interactions related to first service conception in Brangus heifers. Moreover, Lewandowska-Sabat et al. (2012) have integrated genomic and gene expression data aimed at identifying significant pathways involved in the immune response to mastitis in dairy cattle. Here, we have extended our previous genome-wide association study to identify biologically relevant pathways involved in bull fertility. Previously, we have reported eight SNPs associated with SCR with a genome-wide adjusted *P*-value below 0.10 (Peñagaricano et al., 2012). Thus, beyond the enumeration of possible candidate genes (i.e., genes harboring the most significant genetic markers) done in our previous report, in this study we have performed a gene-set enrichment analysis to characterize quantitative trait pathways related to bull fertility. The identified pathways could contribute to a better understanding of the genetic architecture of male fertility in dairy cattle. Moreover, as discussed by Snelling et al. (2012), genetic markers within and surrounding genes initially identified by gene-set enrichment analysis may facilitate the incorporation of marker-assisted selection for complex traits in commercial breeding schemes.

One potential limitation of this study is the limited number of GO categories and IP entries that were analyzed. To avoid testing overly narrow or broad functional categories, while at the same time address the multiple-testing issue, we have analyzed only functional categories with more than 30 genes and, in the case of GO terms, those that were located between levels 5 and 9 in the GO hierarchy. As a result, a total of 662 GO categories and 254 IP entries were interrogated. It is important to note that a key factor in these analyses is the total number of genes, which determines the total number of functional categories that can be inferred. Here, using data from 38,650 SNPs spanning the entire bovine genome, we could analyze a total of 16,819 annotated genes. Future studies, however, could analyze a larger numbers of genes using a high-density SNP chips (e.g., Illumina Bovine HD 770K), which in turn will increase the number of functional categories and provide even further insight into the genetic architecture of bull fertility.

CONCLUSION

In this study, we used genome-wide association SNP results to infer quantitative trait pathways associated with bull fertility. We have identified 20 GO categories and one IP entry that showed significant overrepresentation of genes statistically associated with SCR in Holstein dairy cattle. Interestingly, several of these functional categories, such as small GTPases mediated signal transduction, neurogenesis, calcium ion binding, and cytoskeleton are involved in biological processes closely related to male fertility. These findings could contribute to a better understanding of the genetic architecture of this complex trait in dairy cattle. In addition, pathways enriched with genes that are significantly associated with male fertility could be used in marker-assisted selection in dairy cattle breeding schemes. Thus, the quantitative trait pathway

approach used in this study could complement genome-wide association studies for investigating quantitative traits in livestock species.

ACKNOWLEDGMENTS

The authors thank the USDA-ARS Beltsville Agricultural Research Center (Beltsville, MD, USA) for providing genotypic and phenotypic data. This study was supported by USDA Hatch grant No. 142-PRJ16JH from the University of Wisconsin-Madison.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Livestock_Genomics/10.3389/fgene.2012.00307/abstract

REFERENCES

- Abane, R., and Mezger, V. (2010). Roles of heat shock factors in gametogenesis and development. *FEBS J.* 277, 4150–4172.
- Al-Shahrour, F., Díaz-Uriarte, R., and Dopazo, J. (2004). FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* 20, 578–580.
- Arand, M., Cronin, A., Oesch, F., Mowbray, S. L., and Jones, T. A. (2003). The telltale structures of epoxide hydrolases. *Drug Metab. Rev.* 35, 365–383.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 963–971.
- Björk, J. K., and Sistonen, L. (2010). Regulation of the members of the mammalian heat shock factor family. *FEBS J.* 277, 4126–4139.
- Blaschek, M., Kaya, A., Zwald, N., Memili, E., and Kirkpatrick, B. W. (2011). A whole-genome association analysis of noncompensatory fertility in Holstein bulls. *J. Dairy Sci.* 94, 4695–4699.
- Bragdon, B., Moseychuk, O., Saldanha, S., King, D., Julian, J., and Nohe, A. (2011). Bone morphogenetic proteins: a critical review. *Cell. Signal.* 23, 609–620.
- Braundmeier, A. G., and Miller, D. J. (2001). The search is on: finding accurate molecular markers of male fertility. *J. Dairy Sci.* 84, 1915–1925.
- Brucker, C., and Lipford, G. B. (1995). The human sperm acrosome reaction: physiology and regulatory mechanisms. An update. *Hum. Reprod. Update* 1, 51–62.
- Bustelo, X. R., Sauzeau, V., and Berenjeno, I. M. (2007). GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo. *Bioessays* 29, 356–370.
- Casali, E. A., da Silva, T. R., Gelain, D. P., Kaiser, G. R., Battastini, A. M., Sarkis, J. J., et al. (2001). Ectonucleotidase activities in Sertoli cells from immature rats. *Br. J. Med. Biol. Res.* 34, 1247–1256.
- Chasman, D. I. (2008). On the utility of gene set methods in genomewide association studies of quantitative traits. *Genet. Epidemiol.* 32, 658–668.
- Darszon, A., Labarca, P., Nishigaki, T., and Espinosa, F. (1999). Ion channels in sperm physiology. *Physiol. Rev.* 79, 481–510.
- de Roos, A. P., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–1512.
- DeJannette, J. M., Marshall, C. E., Lenz, R. W., Monke, D. R., Ayars, W. H., and Sattler, C. G. (2004). Sustaining the fertility of artificially inseminated dairy cattle: the role of the artificial insemination industry. *J. Dairy Sci.* 87, E93–E104.
- Fortes, M. R., Reverter, A., Hawken, R. J., Bolormaa, S., and Lehnert, S. A. (2012a). Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. *Biol. Reprod.* 87, 58.
- Fortes, M. R., Snelling, W. M., Reverter, A., Nagaraj, S. H., Lehnert, S. A., Hawken, R. J., et al. (2012b). Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *J. Anim. Sci.* 90, 2894–2906.
- García-Herrero, S., Garrido, N., Martínez-Conejero, J. A., Remohí, J., Pellicer, A., and Meseguer, M. (2010). Ontological evaluation of transcriptional differences between sperm of infertile males and fertile donors using microarray analysis. *J. Assist. Reprod. Genet.* 27, 111–120.
- García-Herrero, S., Garrido, N., Martínez-Conejero, J. A., Remohí, J., Pellicer, A., and Meseguer, M. (2011). Differential transcriptomic profile in spermatozoa achieving pregnancy or not via ICSI. *Reprod. Biomed. Online* 22, 25–36.
- Griswold, M. D. (1998). The central role of Sertoli cells in spermatogenesis. *Semin. Cell Dev. Biol.* 9, 411–416.
- Hadley, M. A., Byers, S. W., Suárez-Quian, C. A., Kleinman, H. K., and Dym, M. (1985). Extracellular matrix regulates Sertoli cell differentiation, testicular cord formation, and germ cell development in vitro. *J. Cell Biol.* 101, 1511–1522.
- Iguchi, N., Tobias, J. W., and Hecht, N. B. (2006). Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7712–7717.
- Iida, H., Yoshinaga, Y., Tanaka, S., Toshimori, K., and Mori, T. (1999). Identification of Rab3A GTPase as an acrosome-associated small GTP-binding protein in rat sperm. *Dev. Biol.* 211, 144–155.
- Khatib, H., Monson, R. L., Huang, W., Khatib, R., Schutzkus, V., Khateeb, H., et al. (2010). Short communication: validation of in vitro fertility genes in a Holstein bull population. *J. Dairy Sci.* 93, 2244–2249.
- Kim, E. S., and Kirkpatrick, B. W. (2009). Linkage disequilibrium in the North American Holstein population. *Anim. Genet.* 40, 279–288.
- Lewandowska-Sabat, A. M., Günther, J., Seyfert, H. M., and Olsaker, I. (2012). Combining quantitative trait loci and heterogeneous microarray data analyses reveals putative candidate pathways affecting mastitis in cattle. *Anim. Genet.* 43, 793–799.
- Li, G., Peñagaricano, F., Weigel, K. A., Zhang, Y., Rosa, G. J. M., and Khatib, H. (2012). Comparative genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. *J. Dairy Sci.* 95, 6122–6129.
- Lie, P. P., Mruk, D. D., Lee, W. M., and Cheng, C. Y. (2010). Cytoskeletal dynamics and spermatogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1581–1592.
- Lucy, M. C. (2001). Reproductive loss in high-producing dairy cattle: where will it end? *J. Dairy Sci.* 84, 1277–1293.
- Lui, W. Y., Lee, W. M., and Cheng, C. Y. (2003). Rho GTPases and spermatogenesis. *Biochim. Biophys. Acta* 1593, 121–129.
- Medina, I., Carbonell, J., Pulido, L., Madeira, S. C., Goetz, S., Conesa, A., et al. (2010). Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res.* 38, 210–213.
- Mruk, D. D., and Cheng, C. Y. (2004). Sertoli-Sertoli and Sertoli-germ cell interactions and their significance in germ cell movement in the seminiferous epithelium during spermatogenesis. *Endocr. Rev.* 25, 747–806.
- Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., et al. (2003). The Interpro database, 2003 brings increased coverage and new features. *Nucleic Acids Res.* 31, 315–318.

- Peñagaricano, F., Weigel, K. A., and Khatib, H. (2012). Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. *Anim. Genet.* 43, 65–71.
- Peng, G., Luo, L., Siu, H., Zhu, Y., Hu, P., Hong, S., et al. (2010). Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur. J. Hum. Genet.* 18, 111–117.
- Pierce, A. A., and Xu, A. W. (2010). De novo neurogenesis in adult hypothalamus as a compensatory mechanism to regulate energy balance. *J. Neurosci.* 30, 723–730.
- Pryce, J. E., Royal, M. D., Garnsworthy, P. C., and Mao, I. L. (2004). Fertility in the high producing dairy cow. *Livest. Prod. Sci.* 86, 125–135.
- R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Snelling, W. M., Cushman, R. A., Fortes, M. R., Reverter, A., Bennett, G. L., Keele, J. W., et al. (2012). Physiology and endocrinology symposium: how single nucleotide polymorphism chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. *J. Anim. Sci.* 90, 1152–1165.
- Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., and Church, G. M. (1999). Systematic determination of genetic network architecture. *Nat. Genet.* 22, 281–285.
- Vazquez, A. I., Bates, D. M., Rosa, G. J. M., Gianola, D., and Weigel, K. A. (2010). Technical note: an R package for fitting generalized linear mixed models in animal breeding. *J. Anim. Sci.* 88, 497–504.
- Wang, K., Li, M., and Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* 8, 1278–1283.
- Weng, L., Macchiardi, F., Subramanian, A., Guffanti, G., Potkin, S. G., Yu, Z., et al. (2011). SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics* 12:99. doi:10.1186/1471-2105-12-9
- Wennemuth, G., Babcock, D. F., and Hille, B. (2003). Calcium clearance mechanisms of mouse sperm. *J. Gen. Physiol.* 122, 115–128.
- Yokota, Y. (2001). Id and development. *Oncogene* 20, 8290–8298.
- Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., et al. (2009). A whole-genome assembly of the domestic cow, *Bos Taurus*. *Genome Biol.* 10, R42.
- Zirkin, B. R. (1998). Spermatogenesis: its regulation by testosterone and FSH. *Semin. Cell Dev. Biol.* 9, 417–421.
- Zöllner, N. (1982). Purine and pyrimidine metabolism. *Proc. Nutr. Soc.* 41, 329–342.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 October 2012; accepted: 11 December 2012; published online: 11 January 2013.

Citation: Peñagaricano F, Weigel KA, Rosa GJM and Khatib H (2013) Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. *Front. Genet.* 3:307. doi: 10.3389/fgenet.2012.00307

This article was submitted to *Frontiers in Livestock Genomics*, a specialty of *Frontiers in Genetics*.

Copyright © 2013 Peñagaricano, Weigel, Rosa and Khatib. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.